# SPECIFICATION

[Electronic Version 1.2.8]

## TITLE OF THE INVENTION

Engineering of Controlled Deamidation Rates in Peptides, Proteins, and Similar Structures

## CROSS-REFERENCE TO RELATED APPLICATIONS

Application # 10/707,263. Design Technique for Use in Engineering of Deamidation Rates of Peptides, Proteins, Hormones, and Peptide-Like, Protein-Like and Hormone-Like Molecules.

## STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

Not Applicable

## INCORPORATION-BY-REFERENCE OF MATERIAL SUMBMITTED ON A COMPACT DISC

Enclosed CD of book: Molecular Clocks: Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins., Robinson, N.E. and Robinson, A.B., Althouse Press, Cave Junction, OR, ISBN 1-59087-250-9. This 448 page book contains a complete review of the subject, including over 1785 references to the research literature, 86 Figures and 16 Tables. The inventions described in this patent are placed in context by this book.

# BACKGROUND OF THE INVENTION

[0001] The deamidation of peptides and proteins as well as molecules related to peptides and proteins is a well known phenomenon. In this reaction, Asn or Gln residues are gradually changed into Asp and Glu residues and their isomers respectively. The rate of this reaction is dependent on the primary sequence, three-dimensional structure, pH, temperature, buffer type, ionic strength and other solution properties. The half-time varies from less than 1 day to more than a century. The reaction introduces a negative charge into the molecule. In addition, the isomerization products ß-Asp and ß-Glu as well D-isomerized forms and chain cleavage also accompany the reaction.

[0002] The stability of Asn and Gln in pharmaceutical and other types of commercial preparations is a major field of study. Efforts have been made to discover formulation conditions that will minimize the rate of deamidation of amides in these preparations. There is also commercial potential in induced or controlled deamidation as an active aspect of the product.

# BRIEF SUMMARY OF THE INVENTION

For the purposes of this work the definition of terms is as follows: Asn - Asparaginyl residue in a peptide or protein, Gln - Glutaminyl residue in a peptide or protein.

[0003] The inventions described here pertain to the engineering of peptides, hormones, and proteins as well as peptide-like, hormone-like and protein-like molecules.

[0004] It is well known that for peptide sequences of the type AsnXxx and GlnXxx, where Xxx is any natural or unnatural amino acid, the rate of deamidation of either Asn or Gln depends very strongly on the identity of Xxx. These results are applicable to peptides, proteins and hormones as well as any amide-containing molecule with similar structure. It is also applicable to isomerization of AspXxx and GluXxx sequences.

[0005] I have done extensive work showing the quantitative sequence dependence of these reactions. I have also invented a method for applying this sequence dependence to proteins, peptides, and other similar molecules, in conjunction with their three-dimensional structures.

[0006] These inventions allow the prediction of deamidation rates of amides as a function of primary and three-dimensional structure, if the three-dimensional

structures is known. They also provide quantitative information about the parameters that make up these rates and show which structural elements are important for each rate.

[0007] These inventions can be used to modify predictably structural elements to provide stability or controlled instability in amides or acids in pharmaceutical and other types of commercial preparations. Specifically there are three major types of modifications that can be made that will change the rate by amounts that can be quantitatively or qualitatively determined from these inventions. Asp and Glu residues also undergo reactions controlled in this way.

[0008] 1. Modification of the residue or residue-like structure to the carboxyl-side or amino-side of the amides or acids. This can be done by substitution of a different natural or non-natural amino acid side chain.

[0009] 2. Exchange of Asn for Gln or Gln for Asn. Gln deamidation and probably Glu isomerization is substantially slower by a quantitative amount.

[0010] 3. Modification of other surrounding structural elements that affect the rate of the reaction as determined by my current three-dimensional calculation procedure or a similar procedure resulting from improvements in the current method.

[0011] These inventions allow the engineering of molecules with specific amide structures that will deamidate at specified rates. These procedures can be used to design stable and unstable forms for pharmaceutical, industrial, and other products. This can be used to increase the shelf-life of such products through minor modifications, prevent or at least slow down the gradual formation of impurities in preparations with these modifications, and may make possible as a result of minor modifications the use of products that would otherwise be too unstable for practical purposes. The engineering of products with unstable amides that are programmed to deamidate at specific rates is also a valuable application of this procedure.

# BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING(S).

Not Applicable

# DETAILED DESCRIPTION OF THE INVENTION

[0012] **General Method:**

[0013] It was known before the invention of the method described here, that the sequence and structure around an amide has a large effect on the deamidation rate. Unknown, however, was the necessary quantitative information that would allow engineering of stable amides or amides with predetermined rates based on modification of the sequence and three-dimensional structure.

[0014] Experiments which I carried out showed quantitatively the effects of sequence dependent deamidation. One of the discoveries made was that the sequence dependence of deamidation is much richer and covers a wider range than previously thought. In 37°C, pH 7.4, 0.15 M Tris buffer, the combination of XxxAsnYyy and XxxGlnYyy sequences where Xxx and Yyy are any of the naturally occuring amino acids covers a range from less than 1 day to over 15,000 days with the entire range in between available.

[0015] In addition to the sequence dependent work, I have also invented methods that allow application of this sequence dependent data to three-dimensional protein structures to permit the prediction of protein deamidation rates. This method is applicable to any peptide type structures including peptides, hormones, and proteins and peptide-like, hormone-like, and protein-like molecules, as well as similar structures that deamidate in the same way.

[0016] This prediction procedure is based on identifying structural elements in a protein or similar molecule that contribute to the rate in known quantitative ways. These include, but are not limited to, hydrogen bonds of various types, disulfide bonds, alpha-helices, and beta-sheets. The effect of each structure depends on a variety of quantitative factors.

[0017] The invention of these prediction techniques had never been attempted before. Not only do they allow prediction of deamidation rates to very high reliability, but the calculation shows what structural features are responsible for each particular rate and what changes would be necessary to modify the rate in a quantitative manner.

[0018] **Sequence Dependence:**

[0019] Tables 1 and 2 show the sequence dependence of deamidation measured using natural amino acid variations in pentapeptides. Non-natural variations provide an even greater range of sequences to choose from.

[0020] Table 1 describes the sequence dependence of Asn sequences. It is based on pentapeptide rates measure in 37°C, pH 7.4, 0.15 M Tris buffer. The applicability of a pentapeptide model to sequence dependence was verified in a separate set of experiments. All values listed in this table are experimental except for the four values in boxes, which were estimated from the rest of the data.

**Table 1 - First-Order Deamidation Halftimes of GlyXxxAsnYyyGly in days at pH 7.4, 37.0 °C, 0.15 M Tris HCl**

| Xxx\Yyy | Gly | His | Ser | Ala | Asp | AmCys | Thr | Cys | Lys | Met | Glu | Arg | Phe | Tyr | Trp | Leu | Val | Ile | Pro | Median† |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gly | **1.03** | **9.2** | **11.8** | **21.1** | **28.0** | **27.6** | **39.8** | **40.6** | **48.2** | **50.4** | **73.9** | **57.8** | **64.0** | **63.6** | **77.1** | **104** | **224** | **287** | **7170** | 50.4 |
| Ser | **0.96** | **9.0** | **15.1** | **24.1** | **30.3** | **41.3** | **45.7** | **60.2** | **55.5** | **54.9** | **59.7** | **59.7** | **52.2** | **64.7** | **76.8** | **110** | **233** | **285** | **7060** | 55.5 |
| Thr | **1.04** | **9.6** | **17.1** | **24.6** | **27.9** | **34.4** | **50.0** | **55.5** | **57.6** | **47.6** | **60.8** | **51.2** | **76.4** | **80.6** | **72.5** | **110** | **237** | **279** | **6290** | 55.5 |
| Cys | **1.14** | **10.8** | **19.0** | **26.4** | **30.6** | 38.3 | **48.7** | **46.0** | **46.6** | **64.5** | **48.3** | **83.1** | **73.9** | **83.9** | **111** | **119** | **229** | **304** | **1550** | 48.7 |
| AmCys | **1.14** | **10.9** | **15.4** | **21.5** | **32.9** | **39.3** | **41.7** | 46.6 | **48.9** | **56.5** | **45** | **58.8** | **63.3** | **78.8** | **81.3** | **100** | **215** | **250** | **3900** | 48.9 |
| Met | **1.04** | **10.2** | **15.2** | **22.1** | **26.4** | **33.8** | **43.6** | **49.6** | **60.4** | **56.9** | **72.4** | **58.8** | **61.9** | **74.0** | **92.7** | **113** | **211** | **275** | **9300** | 57.9 |
| Phe | **1.15** | **10.2** | **18.1** | **24.2** | **27.4** | **29.8** | **39.0** | **46.5** | **58.2** | **58.6** | **62.4** | **61.2** | **69.5** | **75.1** | **102** | **118** | **203** | **287** | **7990** | 58.6 |
| Tyr | **1.49** | **10.2** | **11.9** | **24.3** | **28.4** | **33.3** | **38.1** | **48.6** | **55.1** | **64.3** | **41.0** | **56.9** | **58.0** | **70.6** | **120** | **118** | **241** | **306** | **9830** | 51.8 |
| Asp | **1.53** | **9.7** | **17.0** | **24.0** | **29.4** | **45.8** | **52.4** | **54.1** | **75.9** | **57.3** | **46.8** | **87.2** | **70.1** | **70.4** | **80.3** | **111** | **241** | **298** | **11800** | 55.7 |
| Glu | **1.45** | **9.0** | **16.4** | **25.8** | **32.0** | **32.1** | **36.8** | **44.2** | **77.8** | **59.6** | **60.3** | **80.9** | **70.2** | **94.5** | **98.4** | **130** | **268** | **279** | 9600 | 59.9 |
| His | **1.14** | **10.7** | **15.7** | **24.6** | **31.2** | **33.8** | **47.2** | **43.9** | **50.2** | **63.1** | **69.4** | **48.9** | **72.1** | **82.3** | **95.4** | **116** | **247** | **327** | **8440** | 50.2 |
| Lys | **1.02** | **10.5** | **15.6** | **23.6** | **34.0** | **36.5** | **58.1** | **49.0** | **53.5** | **60.9** | **72.5** | **57.4** | **70.1** | **96.7** | **98.1** | **119** | **246** | **313** | **4940** | 58.1 |
| Arg | **1.00** | **10.0** | **14.3** | **24.4** | **34.7** | **42.3** | **50.7** | **50.5** | **49.6** | **74.4** | **68.3** | **67.4** | **68.3** | **90.0** | **127** | **128** | **247** | **311** | **5790** | 67.4 |
| Ala | **1.05** | **9.3** | **14.9** | **22.5** | **31.9** | **40.6** | **43.5** | **63.7** | **55.9** | **59.2** | **74.1** | **62.4** | **65.6** | **73.9** | **130** | **124** | **254** | **300** | **7370** | 62.4 |
| Leu | **1.08** | **10.7** | **16.7** | **25.1** | **32.1** | **33.6** | **46.1** | **53.5** | **60.1** | **62.6** | **56.7** | **62.1** | **72.4** | **75.7** | **74.5** | **155** | **294** | **391** | **10500** | 60.1 |
| Val | **1.23** | **10.2** | **18.2** | **27.5** | **33.5** | **34.7** | **49.9** | **63.2** | **63.8** | **65.7** | **64.8** | **67.4** | **66.6** | **79.2** | **88.9** | **154** | **291** | **366** | **8030** | 64.8 |
| Ile | **1.26** | **11.5** | **14.5** | **25.9** | **33.8** | **33.0** | **46.3** | **52.7** | **64.4** | **58.8** | **58.6** | **66.4** | **61.5** | **79.3** | **86.7** | **154** | **295** | **384** | **11600** | 58.8 |
| Trp | **1.75** | **11.3** | **15.5** | **30.7** | **43.6** | **42.9** | **38.9** | **83.1** | **59.4** | **64.2** | **75.7** | **73.9** | **71.1** | **92.6** | **135** | **133** | **226** | **286** | 12000 | 67.6 |
| Pro | **1.18** | **12.8** | **18.9** | **31.8** | **48.6** | **43.7** | **63.1** | **60.0** | **67.8** | **78.4** | **92.0** | **72.9** | **100** | **114** | **122** | **181** | **364** | **455** | **6590** | 72.9 |
| Mean | 1.19 | 10.3 | 15.9 | 25.0 | 32.5 | 36.7 | 46.3 | 53.2 | 58.4 | 60.9 | 63.3 | 65.0 | 68.8 | 81.1 | 98 | 126 | 251 | 315 | 7000 | 60.9 |
| St.Dev. | 0.05 | 0.23 | 0.49 | 0.67 | 1.3 | 1.2 | 1.7 | 2.4 | 2.1 | 1.8 | 3.1 | 2.5 | 2.3 | 3.0 | 4.9 | 5.1 | 9.3 | 12.2 | 600 | 2.3 |
| %St.Dev. | 4.4 | 2.2 | 3.1 | 2.7 | 4.1 | 3.3 | 3.6 | 4.5 | 3.6 | 2.9 | 4.8 | 3.9 | 3.4 | 3.7 | 5.0 | 4.0 | 3.7 | 3.9 | 8.8 | 3.7 |
| Median | 1.14 | 10.2 | 15.6 | 24.4 | 31.9 | 34.7 | 46.1 | 50.5 | 57.6 | 59.6 | 62.4 | 62.1 | 69.5 | 79.2 | 95 | 119 | 241 | 300 | 7100 | 59.6 |

† Median does not include Yyy = AmCys

Bold type values are experimental

[0021] Table 2 describes the sequence dependence of Gln peptides. It is also based on pentapeptide rates measure in 37°C, pH 7.4, 0.15 M Tris buffer. In this case, the 52 values shown in bold were measured, and the rest of the values were derived from surface fitting.

[0021b] Tables 1 and 2 were published in: Robinson, N.E., Robinson, Z.W., Robinson, B.R., Robinson, A.L., Robinson, J.A., Robinson, M.L., and Robinson, A.B., (May 2004) Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides, *J. Peptide Res.*, **63**, 426-436.

**Table 2 - First-Order Deamidation Halftimes of GlyXxxGlnYyyGly in days at pH 7.4, 37.0 °C, 0.15 M Tris HCl**

| Xxx\Yyy | Gly | Cys | Met | Thr | Ser | Ala | His | Lys | Leu | Ile | Val | Arg | Glu | Asp | Phe | Pro | Tyr | Trp | Median |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cys | **560** | 800 | 3200 | 3500 | 3800 | 4100 | 4200 | 4400 | 4800 | 4900 | 5000 | 5100 | 5600 | 6100 | 6500 | 7100 | 7900 | 9100 | 4800 |
| Met | 600 | 900 | 3500 | 3800 | 4100 | **4400** | 4400 | 4600 | 5000 | 5000 | 5000 | 5100 | 5800 | 6200 | 6600 | 7300 | 8200 | 9400 | 5000 |
| Thr | **670** | 1000 | 3700 | 4000 | 4200 | 4300 | 4500 | 4800 | 5200 | 5300 | 5100 | **5100** | 5900 | 6300 | 6800 | 7500 | 8400 | 9700 | 5100 |
| Lys | 650 | 1000 | 4000 | 4100 | 4200 | **4300** | **6100** | 4000 | 5300 | 5400 | 5700 | **2300** | **5400** | 5900 | 7000 | 7700 | 8800 | **10000** | 5300 |
| Arg | **660** | **1000** | 4100 | 4200 | 4300 | 4400 | 4900 | 4000 | 5400 | 5500 | 5800 | 2300 | 5400 | 5900 | 7100 | 8100 | 9200 | 11000 | 4900 |
| Val | 640 | 1300 | 4200 | 4300 | 4400 | **4500** | 5000 | **5200** | 5500 | **5600** | 5900 | 6100 | 6500 | 7000 | 7200 | 8500 | 9700 | 12000 | 5500 |
| Pro | 630 | 1600 | 4500 | 4600 | 4600 | **4700** | 5200 | 5500 | 5800 | 6000 | 6200 | **6400** | 6800 | 7200 | 7300 | 8900 | 10000 | 13000 | 5800 |
| Ala | **610** | **1900** | 4400 | **5100** | 5200 | **5300** | 5500 | 5700 | 6100 | **6200** | 6400 | **7200** | 7300 | **7400** | 7500 | 9300 | **10000** | 14000 | 6100 |
| Gly | **650** | 1900 | **4500** | 5200 | **5700** | **5900** | 5900 | 6000 | 6200 | 6300 | **6500** | 7200 | 7300 | 7600 | **7600** | **10000** | 12000 | **15000** | 6200 |
| Leu | **670** | 2000 | 4800 | 5300 | 5800 | 6000 | 6100 | 6100 | 6300 | 6500 | 6800 | 7200 | 7400 | 7800 | 8000 | 10000 | 12000 | 16000 | 6300 |
| Ile | **620** | 2000 | **5100** | 5300 | 5800 | 6200 | 6100 | 6100 | 6300 | 6500 | 7100 | 7200 | 7700 | 8100 | 8100 | 10000 | 12000 | 16000 | 6300 |
| Phe | 660 | 2000 | 5100 | 5300 | 5900 | **6300** | 6200 | 6200 | 6400 | 6400 | 7100 | 7200 | 8100 | 8200 | 8200 | 10000 | 12000 | 16000 | 6400 |
| Ser | 700 | 2100 | 5100 | 5400 | 6000 | **6400** | **6500** | 6300 | 6100 | **5900** | 6800 | 7200 | 8100 | 8200 | 8300 | 10000 | 13000 | 17000 | 6400 |
| Glu | 750 | 2100 | 5200 | 5400 | 6100 | **7100** | 2500 | 4600 | 4300 | **4200** | 6400 | 5200 | 8200 | 8300 | 8400 | 10000 | 13000 | 17000 | 5400 |
| Asp | 800 | 2100 | 5200 | 5400 | 6200 | **7100** | **2500** | **4600** | 6200 | 6400 | 6600 | 5200 | 8200 | 8400 | 8500 | 11000 | 13000 | 17000 | 6200 |
| His | **850** | 2200 | 5200 | 5500 | 6300 | 7200 | 7200 | 4000 | 6600 | 6700 | 6800 | **4500** | 5800 | **5600** | 8600 | 11000 | 14000 | 18000 | 6300 |
| Tyr | **850** | 2200 | 5300 | 5600 | 6400 | 7300 | 7400 | 7500 | 7800 | 7900 | 8000 | 8100 | 8300 | 8600 | 8700 | 11000 | 14000 | **18000** | 7800 |
| Trp | 850 | 2300 | 5300 | **5600** | 6500 | 7400 | 7500 | 7600 | 7900 | 8000 | 8200 | 8300 | 8500 | **8800** | **8600** | 11000 | 14000 | 19000 | 7900 |
| Mean | 690 | 1700 | 4600 | 4900 | 5300 | 5700 | 5400 | 5400 | 6000 | 6000 | 6400 | 5900 | 7000 | 7300 | 7700 | 9400 | 11200 | 14300 | 6000 |
| St.Dev. | 22 | 129 | 163 | 169 | 228 | 296 | 352 | 272 | 226 | 233 | 221 | 423 | 273 | 259 | 180 | 329 | 521 | 809 | 246 |
| %St.Dev. | 3.2 | 7.6 | 3.5 | 3.4 | 4.3 | 5.2 | 6.5 | 5.0 | 3.8 | 3.9 | 3.4 | 7.2 | 3.9 | 3.5 | 2.3 | 3.5 | 4.7 | 5.7 | 4 |
| Median† | 660 | 1950 | 4650 | 5250 | 5750 | 5950 | 6000 | 6050 | 6250 | 6400 | 6650 | 7200 | 7350 | 7700 | 7800 | 10000 | 12000 | 15500 | 6150 |

† Median without charged residues.

Bold type values are experimental.

[0022] Deamidation rates are affected by a wide variety of parameters, including, pH, Temperature, Ionic Strength, and Buffer Ions. These rates are measured under pH and Temperature conditions that are applicable to biological systems. The buffer type and concentration were chosen to minimize ion affects to the extent possible given the experimental limitations. Modification of these conditions will change the rates in Tables 1 and 2. As long as the conditions are not taken to extremes (i.e. high temperature, or strongly acidic or basic conditions) the qualitative sequence dependence should remain the same and the rates reported here can be used with necessary adjustments.

[0023] It is also clear that direct hydrolysis of Gln and Asn take place in addition to the regular sequence dependent mechanism. This hydrolysis is sequence dependent as well, but an average value of about an 8000 day half-time can be taken as a rough approximation based on this and other data measured at the same time. This does not effect the Asn rates significantly, but is responsible for the leveling off of the Gln rates at around this level. This hydrolysis is also effected by the reaction conditions.

[0024] The sequence dependence apparent in Tables 1 and 2 is of great value in engineering stable amides, unstable amides, or amides with particularly desired rates. Isomerization of acid residues will follow a very similar sequence dependence, offset by a determinable amount.

[0025] **Gln vs. Asn Deamidation**:

[0026] It is apparent from the data shown in Tables 1 and 2 that the deamidation rates of Asn and Gln cover markedly different ranges. One of the discoveries in these experiments was that their sequence dependencies are complementary. Asn sequences cover the range from about 1 day to 450 days. Gln picks up at 560 days and carries these rates out to tens of thousands of days.

[0027] This opens up a new possibility for engineering of amide rates. It is possible to switch half-time ranges simply by substituting Asn for Gln or Gln for Asn depending on the desired effect. In many cases where it is desirable to introduce or leave in place an amide, the difference of one $CH_2$ group in chain length may not be critical.

[0028] Moreover, the fact that this range switching can be done raises another possibility. Other modifications of Gln and Asn may lie in different ranges. Thus the substitution of unnatural amide side-chains is also a valuable procedure.

[0029] **Three-Dimensional Effects of Deamidation**:

[0030] The invention of the three-dimensional prediction method for deamidation rates has been developed in two phases. The first of these was the invention of a technique for determining deamidation rates in proteins based on manually counting the number of each type of structure that can affect the rate. Each of these effects is

then summed with special coefficients to produce the correct rate. The procedure was calibrated on known relative deamidation rates and then found to be quite accurate in predicting absolute rates.

[0031] Secondly, the procedure was adapted to an automated method by means of an extensive C++ program. Some modifications were made when this was done, but the basic procedure remained the same.

[0032] I am not attempting to patent this C++ program. There are many ways to write such programs and the current version is protected by copyright. What is being patented is the method used to write it which is based on the manual procedure and minor modifications and improvements that are particularly adapted to computerized calculation and include many conceptual innovations.

[0033] It will be obvious to anyone who studies and understands these methods that there are variations in the procedure and even some improvements that could be made which would yield similar results. Any such modifications are understood to be products of this invention and come under the scope of this patent.

[0034] The deamidation coefficient, $C_D$, for and amide is defined as: $C_D = (0.01)(t_{p1/2})(e^{f(Cm, CSn, Sn)})$

[0035] Here $t_{1/2}$ is the pentapeptide primary structure half life, $C_m$ is a structure proportionality factor, $C_{Sn}$ is the 3D structure coefficient for the nth structure observation, $S_n$ is that observation, and $f(C_m, C_{Sn}, S_n) = C_m[(C_{S1})(S_1) + (C_{S2})(S_2) + (C_{S3})(S_3) - (C_{S4,5})(S_4)/(S_5) + (C_{S6})(S_6) + (C_{S7})(S_7) + (C_{S8})(S_8) + (C_{S9})(S_9) + (C_{S10})(1 - S_{10}) + (C_{S11})(5 - S_{11}) + (C_{S12})(5 - S_{12})]$. The structure observations, $S_n$, were selected as those most likely to impede deamidations, including hydrogen bonds, α-helices, β-sheets, and peptide inflexibilities. The functional form of $C_D$ assumes that each of these structural factors is added to the reaction activation energy. The observed $S_n$ were:

[0036] For Asn in an α-helical region:

[0037] $S_1$ = distance in residues inside the α-helix from the $NH_2$ end, where $S_1$ = 1 designates the end residue in the helix, 2 is the second residue, and 3 is the third. If the position is 4 or greater, $S_1 = 0$.

[0038] $S_2$ = distance in residues inside the α-helix from the COOH end, where $S_1 = 1$ designates the end residue in the helix, 2 is the second residue, and 3 is the third. If the position is 4 or greater or $S_1 \neq 0$, then $S_2 = 0$.

[0039] $S_3 = 1$ if Asn is designated as completely inside the α-helix, because it is 4 or more residues from both ends. If the Asn is completely inside, $S_3 = 1$, $S_1 = 0$, and $S_2 = 0$. If $S_1 \neq 0$ or $S_2 \neq 0$, then $S_3 = 0$.

[0040] For flexibility of a loop including Asn between two adjacent antiparallel βsheets:

[0041] $S_4$ = number of residues in the loop.

[0042] $S_5$ = number of hydrogen bonds in the loop. $S_5 \geq 1$ by definition.

[0043] For hydrogen bonds:

[0044] $S_6$ = the number of hydrogen bonds to the Asn side chain C=O group. Acceptable values are 0, 1, and 2.

[0045] $S_7$ = the number of hydrogen bonds to the Asn side chain $NH_2$ group. Acceptable values are 0, 1, and 2.

[0046] $S_8$ = the number of hydrogen bonds to the backbone nitrogen atom in the peptide bond on the COOH side of Asn. Hydrogen bonds counted in $S_6$ or $S_7$ are not included. Acceptable values are 0 and 1. This nitrogen atom is used in the five-membered succinimide ring.

[0047] $S_9$ = additional hydrogen bonds, not included in $S_6$, $S_7$, and $S_8$, that would need to be broken to form the succinimide ring.

[0048] For Asn situated so that no α-helix, β-sheet, or disulfide bridge structure is between the Asn and the end of the peptide chain:

[0049] $S_{10}$ = 1 if the number of residues between the Asn and the nearest such structure is 3 or more. If the number of intervening residues is 2, 1, or 0, or Asn not between structure and chain end, then $S_{10}$ = 0.

[0050] If the Asn lies near to any α-helix, β-sheet, or disulfide bridge structures:

[0051] $S_{11}$ = the number of residues between the Asn and the structure on the $NH_2$ side, up to a maximum of 5. Values of 0, 1, 2, 3, 4, and 5 are acceptable.

[0052] $S_{12}$ = the number of residues between the Asn and the structure on the COOH side, up to a maximum of 5. Values of 0, 1, 2, 3, 4, and 5 are acceptable.

[0053] Hydrogen bonds selected by the Swiss Protein Data Bank (PDB) viewer were accepted if the bond length was 3.3 Å or less and there was room in the structure to accommodate the van der Waals radius of the hydrogen. In the computerized procedure this bond length was optimized at 4.1 Å, and the bond angles and number of bonds per atom were adjusted to physically correct and optimized values. The Swiss PDB viewer, according to the customary criteria, selected α-helices and β-sheets. All primary

structure $t_{1/2}$ values were those published[6], except for Asn with carboxyl-side Pro, Asn, or Gln and N-glycosylated Asn. Estimated values were used for any sequence for which the primary sequence rate was not known.

[0054] **Coefficients Used in Equation:**

[0055] $C_D$ values ("Coefficient of Deamidation") were optimized by using various values for $C_m$ and $C_{Sn}$ to maximize the value of the deamidation resolving power, $D_P$, as described in the calibration procedure section. The optimized values were $C_m = 0.48$, $C_{S1} = 1.0$, $C_{S2} = 2.5$, $C_{S3} = 10.0$, $C_{S4,5} = 0.5$, $C_{S6} = 1.0$, $C_{S7} = 1.0$, $C_{S8} = 3.0$, $C_{S9} = 2.0$, $C_{S10} = 2.0$, $C_{S11} = 0.2$, and $C_{S12} = 0.7$.

[0056] As an example, the β-LysAsn(145)His sequence of hemoglobin is not in an α-helix or in a loop between two βsheets, so $S_1$ through $S_4 = 0$, $S_5 = 1$. There is one hydrogen bond to the amide side chain nitrogen and one other to be broken to form the imide, but there are none to the amide carboxyl or the backbone nitrogen, so $S_6 = 0$, $S_7 = 1$, $S_8 = 0$, and $S_9 = 1$. This Asn is near the carboxyl end of the chain and one residue from an α-helix on the amino side, so $S_{10} = 0$, $S_{11} = 1$, and $S_{12} = 5$. The GlyLysAsnHisGly half life[6] is 10.5 days. Therefore, $C_D = (0.01)(10.5)e^{-(0.48)[(1)(1)+(2)(1)+(2)(10)+(0.2)(4)]} = (0.105)e^{(0.48)(5.8)} = (0.105)(16.184) = 1.70$.

[0057] $C_D$ is multiplied by 100 to give the predicted Tris deamidation half-time in days for the amide.

[0058] Results for Asn are greater than 95% correct in predicting the fastest amide in a protein. It is also applicable to Gln.

[0059] It is also likely that isomerization of Asp and Glu can be modeled with the same procedure. Primary rate data on Asp and Glu isomerization or a correction factor to be applied to the Asn and Gln data is needed in order to do this.

[0060] **Conclusions:**

[0061] Three different types of modifications that can be used in the engineering of deamidation and/or isomerization rates of amides and possibly acids have been invented. These are:

[0062] 1. Modification of the residues or residue-like structures on either side of the amide - principally the one on the right (carboxyl side).

[0063] 2. Modification of the amide - specifically Asn to Gln or Gln to Asn, but other types of modification can also be used, especially in the case of structures that are similar, but not a perfect match to those found in peptides, hormones, and proteins.

[0064] 3. Modification of the three-dimensional environment around the amide. The necessary modifications can be determined from the three-dimensional deamidation prediction method. Each of the S parameters describes a quantitative addition to the reaction activation energy. Removal or addition of one or more of these elements will change the rate accordingly.

[0065] At least two types of deamidation are present. The ones on which this method is based, and which are most prevalent for amides with half-times less than a few hundred days, depending on conditions and providing especially catalytic ions are not present, are most strongly effected by the structure to the right of the amide (e.g. in the sequence GlyXxx(Amide/Acid)YyyGly the identity of Yyy is the most important factor). Also present is at least one more mechanism that is usually slower and has different sequence dependence. It is possible that this dependence as well as the left hand structure dependence (Xxx in the sequence GlyXxx(Amide/Acid)YyyGly) can also be modeled with a similar system, but this has not yet been demonstrated.